

ROI VON DATENQUALITÄTSMASS- NAHMEN

Human Inference - a neopost company



Wann lohnen sich

Investitionen in die Datenqualität?

Der Erfolg von vielen Geschäftsprozessen steht in direktem Verhältnis zur Qualität der Kundendaten. Ob unvollständige oder falsche Eingaben von personenbezogenen Informationen in die unternehmenseigenen Datenbanken, veraltete Angaben, ungewollt mehrfache Aufnahme oder mögliche Dubletten: mangelnde Effizienz in Marketing und Vertrieb, verärgerte Kunden oder erschwertes Neukundengeschäft sind fast immer die Folge.

Mittlerweile gibt es einige Verfahren, um die "verschmutzten Daten" zu bereinigen und auf den aktuellen und korrekten Stand zu bringen. Doch wie kann man ermitteln, ob sich entsprechende Investitionen – zum Beispiel für Data Profiling, Data Cleansing oder Monitoring – wirklich lohnen? Bis heute haben sich in Wissenschaft und Praxis für viele Datenqualitätsmerkmale noch keine Verfahren durchgesetzt, die eine gezielte Quantifizierung der Datenqualität erlauben. So ist es zwar vergleichsweise einfach, auf der Basis retournierter Briefe die unnötigen Portokosten zu eruieren, die wegen fehlerhafter Adressdaten in der Kundendatenbank entstanden sind. Wesentlicher schwieriger stellt sich allerdings dar, den Imageverlust zu beziffern, wenn ein Kunde aufgrund bestehender Dubletten im Datenbestand das gleiche Schreiben mehrfach erhält, deswegen verärgert ist und womöglich zur Konkurrenz wechselt. In diesem Fall resultieren aus der **mangelhaften Datenqualität** letztlich nachhaltige **Umsatzeinbußen** sowie ein geringerer "Customer Lifetime Value" (Kundenwert aus Unternehmenssicht).

Die so entstehenden Kosten fallen dabei in der Regel um einiges höher aus als diejenigen, die wegen fehlerhafter Adressen oder Dubletten mit dem Postversand einhergehen. Ein einfaches **Beispiel**: Ein Schaden in Höhe von 5.000 Euro, der bei einem Mailing durch nicht zustellbare Briefe und Rückläufer entsteht, kann ohne Weiteres Folgen in Millionenhöhe haben – wenn man den entgangenen Umsatz durch nicht erreichte Kunden sowie die Auswirkungen auf den "Customer Lifetime Value" in die Berechnung mit einbezieht.



Qualitätsmerkmale mittels Metriken messen

Für die Planung von Datenqualitätsmaßnahmen lassen sich jedoch inzwischen Metriken festlegen, mit deren Hilfe der aktuelle Stand der Datenqualität gemessen werden kann. Auf dieser Grundlage können entsprechende **Berechnungen des Return on Investment (ROI)** durchgeführt werden, um Entscheidungen über entsprechende Maßnahmen unter Kosten-Nutzen-Gesichtspunkten zu treffen. Im Zentrum steht dabei zunächst die Frage, inwieweit sich die Datenqualität mittels Metriken für einzelne Datenqualitätsmerkmale messen lässt? Danach wird im zweiten Schritt betrachtet, inwieweit die Metrikergebnisse (und damit das Datenqualitätsniveau) durch die Anwendung von Datenqualitätsmaßnahmen verbessert werden können? Als Beispiel soll im Folgenden kurz eine Metrik für das Datenqualitätsmerkmal "Aktualität" erläutert werden:

Dabei wird unter "Aktualität" die Eigenschaft der Gegenwartsbezogenheit des Datenbestandes verstanden – das heißt, inwieweit ein im System vorhandener Datensatz noch den aktuellen Gegebenheiten entspricht und nicht veraltet ist. Im Gegensatz zum Datenqualitätsmerkmal "Fehlerfreiheit" beruht die Überprüfung in diesem Fall auf **wahrscheinlichkeitstheoretischen Betrachtungen**, um eine automatisierte Messung zu ermöglichen. "Aktualität" und damit auch die resultierenden Metrikergebnisse können in diesem Zusammenhang als Wahrscheinlichkeit dafür interpretiert werden, dass der untersuchte Datenwert tatsächlich noch

aktuell ist. Hierbei nimmt in der Regel die Wahrscheinlichkeit, mit der ein bestimmter Wert im Datenbestand zum Analysezeitpunkt noch aktuell ist, mit zunehmendem Alter des Datenwerts ab.

Mathematisch lässt sich dies folgendermaßen ausdrücken: Sei A ein Attribut, w ein Attributwert im Informationssystem und $\text{Alter}(w, A)$ das Alter des Attributwerts, das sich aus dem Zeitpunkt der Messung und dem Zeitpunkt der Datenerfassung errechnet. Sei zudem $\text{Verfall}(A)$ die Verfallsrate von Werten des Attributs A (diese gibt an, wie viele Datenwerte des Attributs durchschnittlich innerhalb einer Zeiteinheit inaktuell werden). Dann ist die Metrik auf Attributwertebenen bei Zugrundelegung einer Exponentialverteilung beispielsweise wie folgt definiert:

$$Q_{\text{Akt.}}(w, A) := e^{-\text{Verfall}(A) * \text{Alter}(w, A)}$$

Aktualität als Erfolgsfaktor

Praktische Anwendung fand die vorgeschlagene Metrik im Rahmen des **Kampagnenmanagements eines großen Mobilfunkanbieters**. Hier war aufgrund mangelhafter Daten oftmals keine korrekte und individuelle Kundenansprache möglich. Dies schlug sich gerade bei Mailingaktionen in geringen Erfolgs- und Abschlussquoten nieder. Deshalb wurde eine geplante Kampagne zur Vermarktung einer Tarifoption unter Verwendung der Datenqualitätsmetrik zunächst hinsichtlich des ROI analysiert und anschließend zielgerichtet und unter Berücksichtigung von Datenqualitätsaspekten durchgeführt. Dabei wurde zunächst für die betroffenen Attribute "Name", "Vorname", "Adresse" und "Produkt" die relative Wichtigkeit im Rahmen der Kampagne bestimmt. Von besonders hoher Wichtigkeit war etwa der aktuelle Tarif des Kunden (Attribut "Produkt"), da nur Kunden mit einem bestimmten Tarif angesprochen werden sollten. Demgegenüber wurde beispielsweise dem (korrekten) Vornamen eine geringere Wichtigkeit zugemessen. Danach erfolgte

für alle betrachteten Attributwerte die Berechnung des Alters $\text{Alter}(w, A)$ aus dem gegenwärtigen Zeitpunkt und dem Zeitpunkt der Datenerfassung. Anschließend konnte dann, basierend auf empirisch ermittelten Verfallsparametern $\text{Verfall}(A)$, für jeden Attributwert automatisiert das Metrikergebnis für Aktualität berechnet werden. Die Analyse einer früheren, vergleichbaren Kampagne erlaubte hierbei die Prognose der Profitabilität der neuen Kampagne in Abhängigkeit vom Metrikergebnis für Aktualität. Die Berechnungen zeigten hierbei, dass die **Erfolgsquote** umso höher ist, je höher die Aktualität der Kundenattribute ist. Auch ließ sich ableiten, dass es im Rahmen der Kampagne unter Kosten-Nutzen-Gesichtspunkten nicht sinnvoll war, Kunden mit einem geringen Metrik-

ergebnis für Aktualität überhaupt erst anzuschreiben.

Des Weiteren konnten auf Basis der Metrikergebnisse Business Cases

für potenzielle Daten-

qualitätsmaßnahmen aufgestellt werden. So ergab eine Analyse beispielsweise, dass der Adresszukauf nur für ausgewählte Datensätze (in Abhängigkeit vom Metrikergebnis) ökonomisch sinnvoll ist. Insgesamt zeigt das Beispiel, dass sich der ROI für Maßnahmen zur Verbesserung der Datenqualität durchaus fundiert und nachprüfbar ermitteln lässt. Zudem bleibt festzuhalten, dass Investitionen in die Datenqualität neben Kosteneinsparungen auch eine Verbesserung der Wertschöpfung zur Folge haben sowie neue Geschäftschancen eröffnen können. Aktuelle und korrekte Kundendaten sind hierbei nicht nur für eine erfolgreiche Segmentierung und Kundenansprache erforderlich.

Vielmehr ist das richtige Wissen über die eigenen Kunden auch eine unverzichtbare Grundlage für die Entwicklung neuer Produkte und Services. Die Kombination aus alledem kann wiederum die Basis dafür bilden, nachhaltig Wettbewerbsvorteile zu generieren und zu realisieren.